The safety management of urban rail transit based on operation fault log

Xiaobing Ding *, Xuechen Yang, Hua Hu, Zhigang Liu

School of Urban Rail Transportation, Shanghai University of Engineering Science, Shanghai 201620, China

Article history:
Received 23 August 2016
Received in revised form 18 December 2016
Accepted 19 December 2016

Keywords:
Safety management
Metro safety
Operation fault log
Data mining
Database technology

Abstract
With the rapid development of China's rail transit, the safety of Metro has roused the society's concern more and more widespread. So deeply mining the massive dispatching log data is of great significance to the safety management of Metro operation. For the purposes of risk early-warning and promoting the safety management level of Metro, a dispatching fault log management and analysis database system (DFLMIS) is designed, which contains almost all kinds of accidents that have occurred in the operation of Metro. Taking the compatibility and safety into consideration, the Visual studio 2010 and SQL-sever 2005 are used to develop the DFLMIS. First, changing operation fault log is regarded as a state machine, which describes the data from three dimensions: time, value of information, frequency, and forms the operation scheduling database with data management as the visual angle. Second, the probability space cut algorithm is presented for pruning strategy of probability space, which is suitable for high frequent update of the environment of grid technology as index structure. Finally, the procedures are demonstrated to how DFLMIS can be used to early-warn and identify the risk sources. The research and design of DFLMIS would be of great help to the Metro operators to identify the risk and promote the safety management level.

1. Introduction

During the Nearly 10 years, China's urban rail transit line operating mileage grows faster. Up to the end of 2015, 33 cities have been approved or been operating metro, and the total amount of operation lines has reached 96, the total operating mileage has reached 4000 km. According to the plan of line network, by 2020, the total operating mileage of urban rail traffic is expected to break through the 7200 km, the network trend is very obvious. The passengers carried by metro continue to increase, and the complexity of operation organization increases. The factors affecting the operation safety and reliability become more, fault rate continues increasing. Therefore, how to improve the reliability of rail transit hardware equipment, passenger service, reduce the occurrence of fault is very important.

The main accidents caused by Metro safety both in domestic and international have become increasingly prominent: The South Korea Daegu subway fire happened on February 18, 2003, which caused 198 peoples' death and 146 peoples' injury, and resulted in property losses up to 4 billion 700 million won (Balducelli and D'Esposito, 2000).

On December 22, 2009, the crash of Shanghai train leads to the surrounding road traffic paralyzed, and it cost 105 buses 6 h to evacuate stagnation passengers; On September 27, 2011, more than 260 people were injured in Shanghai train crash accident (Kyriakidis et al., 2012). These series of accidents in the world caused strong social repercussions, how to efficiently disposal emergencies during the urban rail traffic incidents, has widely aroused the extensive concern of the whole society and universal attention. The safety management of urban rail transit operation has become an urgent problem to be solved. The basic procedure of data mine can be illustrated in Fig. 1.

2. Literature review

2.1. Operation data of rail transit systems

The operating data of subway contains line operating mileage, train interval running time, operation failure, passenger flow, travel OD data, etc., the safety related data of subway is not only less, but also difficult to obtain. After many years' operation, Beijing, Shanghai and other cities, which operate the subway earlier, has accumulated a large number of operational data (Shi, 2004), domestic and foreign scholars have made a more basic research based on these operational data about safety management of subway.

He (2015) the information construction of urban rail transit enterprise is often a single line sorted by professional, and lack of overall planning, coupled with the construction of large span,
information standards not unified, which will cause to be lack of an effective linkage between each line, information isolated island is formed, resulted in each subject cannot be timely communication and coordination, which will give the overall efficiency of the operation and management bring negative influence, and even affect the efficiency of the disposal of emergency. Dobbins (2011) designed an information system to support real-time response as well as planning decisions such as risk resource allocation and evaluation of potential response strategies. In the event of an incident, this system enables en route responders to view incident details via an Internet GIS map service. Zhang (2008) the system adopts and integrates the technology of computer vision, digital image processing technology, database technology, communications technology, control technology, image sensor technology and system integration technology to the transportation industry. And using the most advanced technology of union database currently to give the sharing of information resources from different IT application system of transportation industry. The advanced scheme shall become a base for the development of integrated information system for urban transportation (Lu et al., 2011; Reinach and Viale, 2006; Munder and Gavras, 2006).

Form the above, many information systems have been used to analyze and prevent the fault of operation. So it is of great necessity to do research of information system based on the operation log.

2.2. The analysis of operation data

Zhang (2016) an adaptable metro operation incident database is presented for containing details of all incidents that have occurred in metro operation. Microsoft Access 2010 is used for the comprehensive and thorough design of the MOID. Through the statistical characteristics of incident, such as types, causes, time, and severity, and 24 accident precursors are identified from Shanghai metro. Finally an organizational structure is proposed from the four aspects of supervision, research, implementation, and manufacturer. It would be conducive to safety risk analysis in identifying relevant precursors in safety management. Takeda (2015) an extraction method of experts' operation know-how from historical operation data is proposed, and efficiencies of the proposed method are demonstrated by numerical experiments using a dynamic simulator. Wang (2009) the flight operation control risk assessment index is established (Zhang, 2016). Using the Bayesian network analysis method based on the fault tree and insecurity event analysis reports over the years as a sample; the probability of occurrence of unsafe events by forward reasoning is predicted. Risk inference result is entirely consistent with actual operation situation.

2.3. Information system design and application

Jiao (2013) the design and implementation of the Beijing subway risk management platform is discussed, the risk management status and existing problems are combined. After the investigation and research of the risk management of Beijing subway, and then the demand of performance and data analysis platform is analyzed from the aspects of function, according to the system demand analysis, the risk management platform functional design, interface design and database design (Jianjun and Siji, 2002). Thirdly, the author describes the risk management platform of software and hardware architecture.

3. Operation scheduling data processing and mining

The quantity of original operation scheduling log is up to 110,284, which is from Shanghai Metro 6, 9, 10 line arrange from 2011 to 2015, containing all kinds of faults that encountered in the course of rail transit operations, the normal operation of the record, the construction inspection, night exercises and other production activities related to the subway operation (Peng and Wang, 2001; Xu et al., 2009). The normal log records, fault records, redundant data are all included, so it needs data mine and processing to the basic raw data. The hazard data processing procedures can be illustrated in Fig. 2.

3.1. The original data

Shanghai Shentong Metro Dispatch log is recorded by the control center of each line, the dispatching log reflects the non-normal situation during the process of operation, covering: the name of the line, scheduling date, scheduling time, recording content, starting station, terminal station, site type, scheduling, professional record types, equipment number subordinate units, equipment, equipment type, warranty time, registration number, report, fault repair time, responsibility type, reason, reason segmentation, two minutes late to start, two minutes late to arrive, five minutes late to start, five minutes late to arrive, and clear off the table, through the exchange dropped, one of the core fields “content” is the subjective descriptive content (Jin and Sendhoff, 2008). Because of the confidentiality of the data and the length of the space, only parts of the content are chosen from the traffic scheduling log.

In recent years, Shanghai urban rail transit line traffic continues to increase, the impact of operational safety and reliability of the fault factors continue to increase, resulting in a growing trend of failure frequency (Liu and Sun, 2009). Among them, the fault caused by vehicles, general and objective factors the proportion is particularly prominent. Therefore, it is very important to improve the reliability of rail transit hardware and passenger service and reduce the occurrence of faults.

According to the provided operational data from the operators, main fault includes seven categories: vehicles, power supply, communication, maintenance management, passengers’ transportation, and total harmonic, objective causes. It is usually divided the 7 categories into three grades: the first class classification, the second and third classification, the three categories can be subdivided into 81 types of fault.

3.1.1. Fault type 1: Vehicle fault

The fault of the vehicle is divided into: control system, train doors, braking system, traction system, auxiliary system and two
other class classifications, the specific segments including: electrical control equipment, display fault and other control systems, security doors, parking brake, air cylinder, air compressor, BC/B9 valve, other aspects of brake and main circuits, over-speed switching, pantograph, auxiliary inverter, lighting, battery, air conditioner, broadcast, coupler, train body, other three level classification, each type also contains different keywords.

### 3.1.2. Fault type 2: Communication signal fault

As the nervous system of the rail transit network, the communication signal takes on the important responsibility, because of its complexity, the frequency of fault is relatively high.

Communication signal fault is divided into: Vehicle ATC, ATS, trackside equipment, switches, fault of signal, communications and other secondary classification, the specific segments: crashes of VOBC, lost of patterns, lost of position, the red light appearance of track, signal failure, beacon, the network storm, DT fault/communication interrupt and central/station, telephone, wireless intercom, etc.

### 3.1.3. Fault type 3: Power supply fault

The power failure is divided into three grade two classifications: transformer substation, contact line, power supply. Shanghai rail transit takes the two power supplies, the frequency of power supply fault is low, so the power supply part of the probability of failure is relatively small during procession of operation.

### 3.1.4. Fault type 4: Maintenance fault

The maintenance fault can be divided into: line, switches fault and others. The construction fault caused by the daily operation of urban rail transit is relatively small, which is mainly concentrated in the night repair or normal construction inspection. And maintenance scheduling log records information are usually normal operation, the noise data in the data process has been cleaned.

### 3.1.5. Fault type 5: The chief dispatch fault

The chief dispatch fault is the sub-classification of the dispatching order. The main task of the chief dispatch is to control coordinated operation of overall adjustment for the entire rail transit network of the various lines, to ensure that the urban rail transit train lines are of efficient, safe and orderly, the probability failure is very low from the whole macroscopic layer of the rail transit operation control.

### 3.1.6. Fault type 6: Passenger transport related fault

Passenger transport fault can be divided into: station attendant, driver, shielding door/security door, running/signal delays and other two categories of passenger transport. Passenger transport is the primary sector of the passenger service, the frequency of passenger transport is relatively high. Different passenger rail transport operators have different levels of demands; they bear a greater pressure, which also lead to a high frequency of passenger transport sector, high density of the working environment, as a result of the increase in the number of passenger failures occurred relatively to other departments, which is more prominent.

### 3.1.7. Fault type 7: Objective fault

Objective failure can be divided into: crowds, weather, Human-vehicle conflict, foreign object penetration limit, personnel restrictions, passengers, people or goods caught in the doors, confirm the safety conditions, and others. There are many factors leading to this type of fault, the fault frequency is relatively high, because there are many unpredictable factors, which also cannot be prevented.

The seven types of fault occur with different frequency in daily operation, so how to deal with various types of operational failure is a very important part of urban rail transit operation process. Developing effective contingency plans is helpful to operate the dispatch department in emergency situation, more rapid and accurate measures to improve the efficiency of emergency disposal. Therefore, it is of great significance to divide the massive fault data properly (see Fig. 3).

First, the original data is cleaned; then, screen all the data, so achieve the result that the item of hazard source data can be up to 29.7%.

The operation log, which is not related to the accident, is assumed as the noise data, and the noise data is cleaned and deleted according to the contents of the hazard source data. Finally, 21,156 risk source data are obtained. The cleaning results are shown as follows (see Table 1).

### 3.2. Data mining

After the pre-process of operational fault scheduling log, it is of accordance with the conditions of data mining, the Apriori algorithms will be proposed to use to analysis the association rules, the main steps are as follows:

Step1: Initialize database, assume the pre-processed the fault dispatch log as the original data and initialize the database, Scan transaction database $T_D$, Find out all the items in the collection, the length of which is $k = 1$, then candidate set 1 is formed called $C_1$, and the support of each item can be obtained, if it is greater than $\minsupport$, frequent set 1 can be formed, called $L_1$;

Step2: according to frequent set $k$, generate the $k + 1$ candidate set $C_{k+1}$, if $C_{k+1} \neq \phi$, go to the next step, otherwise, end the cycle;

Step3: Scan the database, so as to determine the support degrees of each candidate set;

Step4: Delete candidate items, the support degree of which is less than $\minsupport$, so as to form the $k + 1$ frequent set $L_{k+1}$;

Step5: $k = k + 1$, go to step 2, until all items are traversed;

Step6: The strong association rules can be obtained from the course.

### 4. Dispatching fault log analysis and management database system

In general, every system should contain several functions, in which announcement module, safety check list, hazard source, fault tracking, hazard location, and emergency assistance.

#### 4.1. Data flow chart

As far as the information system is concerned, the relation among the functions of system, database, and data interface is very essential. The database is the data center of the system, all the data used is from it, and it can provide the data for operation risk source module, equipment risk source module, risk source module. From these database modules, high frequency risk can be obtained, and the classification of risk matrix can be made. Finally, the operational and environment risk analysis can be made. Besides, the security control of database and the control policy of risk source
are included. According to the requirements of the rail transit dispatch log management system design and function, system data flow relationships can be shown in Fig. 4.

### 4.2. Relation between tables

Through the analysis of safety management of Metro operations, combined with the modules that the DFLMIS required, there are 10 tables: admin table, main table, dingwei table, risk table, dwpic table, xinxi table, daping table, news table, guzhang table, falv table, so E-R model of the system can be obtained, as is shown in Fig. 5.

### 4.3. Basic functions

In general, the basic functions of DFLMIS can be shown in Fig. 6, every module has its importance during the risk sources management in the course of operation.

The main functional modules include: Announcement, safety check list, hazard source, fault tracking, hazard location, and emergency assistance. Announcement module shows the newest news

---

**Table 1**

<table>
<thead>
<tr>
<th>Data type</th>
<th>Amount of risk source</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Train collision</td>
<td>157</td>
<td>0.74%</td>
</tr>
<tr>
<td>Vehicle derailment</td>
<td>4257</td>
<td>20.12%</td>
</tr>
<tr>
<td>Living space degradation</td>
<td>985</td>
<td>4.66%</td>
</tr>
<tr>
<td>Explosion</td>
<td>7</td>
<td>0.03%</td>
</tr>
<tr>
<td>Fire</td>
<td>54</td>
<td>0.26%</td>
</tr>
<tr>
<td>Emergent events</td>
<td>2473</td>
<td>11.69%</td>
</tr>
<tr>
<td>Fear</td>
<td>3027</td>
<td>14.31%</td>
</tr>
<tr>
<td>Electric shock</td>
<td>2147</td>
<td>10.15%</td>
</tr>
<tr>
<td>Bruise</td>
<td>21</td>
<td>0.10%</td>
</tr>
<tr>
<td>Fall down</td>
<td>5014</td>
<td>23.70%</td>
</tr>
<tr>
<td>Natural environment</td>
<td>3014</td>
<td>14.25%</td>
</tr>
</tbody>
</table>

---

**Fig. 3.** The process flow of metro dispatch log.

**Fig. 4.** Data flow relation of system.
and incident about the urban rail operator, so others can read them in time; the safety check list module helps to quickly search the risk data source by discipline, and learn the emergence plans; all risk source data can be searched from the hazard source module; Fault tracing module can reflect the state of failure timely, in which the red dots indicate fault, while green ones indicate normal status, which can be shown in Fig. 7.

5. Using the system to improve Metro safety

5.1. Collection and analysis of the fault log characteristics

Through nearly 5 years of Shanghai rail transit operation data collection, a total of 11 million operation and dispatching data is collected, which contains part of the normal logging data, after preliminary screening and classification, relevant operational experts are organized to audit the risk source records, and the final risk source data is determined, as the basic input data for the database to form data initialization. Finally, a total of 64,226 fault log data are obtained, and the data characteristics are shown in Fig. 8.

5.2. Hazardous degree analysis

Through the analysis and calculation of the Apriori algorithm and the matlab calculation tool, frequent item sets L can be concluded, because of the large number of frequent item sets (Chen and Yao, 2010), only part of the frequent item sets are listed here, which is shown in Table 2.

Because of the amount of frequent sets, the association rules generated is very large, there is no necessary to be demonstrated by chart expression, according to the income of each association...
rule to calculate the degree of confidence, division grade of frequency fault can be obtained. Fault frequency level can be divided into ten types with the proportion of 5%, and the interval division is 60–100% (see Table 3).

6. Discussion and conclusions

This paper presents a metro operation information system comprising 7 types of fault incidents, near misses, demonstrates its use in accident prevention. Of particular importance is the consideration of near misses, as their data provides a crucial foundation for safety analysis and evaluation. Implementing the DFLMIS will

<table>
<thead>
<tr>
<th>Item sets</th>
<th>Count of support</th>
</tr>
</thead>
</table>

Table 3

<table>
<thead>
<tr>
<th>Division grade of frequency fault.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Range of confidence</td>
</tr>
<tr>
<td>----------------------</td>
</tr>
<tr>
<td>100%</td>
</tr>
<tr>
<td>100–95%</td>
</tr>
<tr>
<td>95–90%</td>
</tr>
<tr>
<td>85–90%</td>
</tr>
<tr>
<td>80–95%</td>
</tr>
<tr>
<td>75–80%</td>
</tr>
<tr>
<td>70–75%</td>
</tr>
<tr>
<td>65–70%</td>
</tr>
<tr>
<td>60–65%</td>
</tr>
<tr>
<td>60%以下</td>
</tr>
</tbody>
</table>

Fig. 7. The main interface of DFLMIS.

Fig. 8. The characteristics of risk source.

Table 2

Frequent item sets L (part).

<table>
<thead>
<tr>
<th>Item sets</th>
<th>Count of support</th>
</tr>
</thead>
</table>
also involve cooperating with and subsidizing a research institution to support their input.

This could be in the form of a management consultancy to help the metro operation company with DFLMIS analysis and management and improving operation safety. The metro operation company itself would play a key role in implementation. This could involve a DFLMIS management committee responsible for establishing the organizational structure and relevant regulations, responding to the management requirements of the government departments, cooperating with the research institution to operate the DFLMIS effectively, and periodic evaluation of its operation. As the Shanghai incident statistics highlight, most incidents are caused by equipment failure, with about 70% of accidents being directly attributable to this cause. The equipment manufacturer, therefore, could use the DFLMIS to analyze the failures that are occurring and help identify their root causes to improve technical design standards and manufacturing processes.

Acknowledgments

The research described in this paper was supported by the National Natural Science Foundation of China (Grant No. 71601110). The authors also gratefully acknowledge those who provided data and suggestions.

References